ANALYSIS OF THE RELATIONSHIP BETWEEN FACILITY CHARACTERISTICS AND OIL SPILL RISK

1. EXECUTIVE SUMMARY

This analysis investigates whether relationships exist between the characteristics of oil storage facilities and spill incidents. This analysis is based on the results of EPA's 1995 Survey of Oil Storage Facilities (1995 SPCC Survey), which collected information from more than 2,600 oil-storing facilities in 23 different industries, and provided information on facility and tank characteristics, oil spill incidents, and facility operations. The results of the analysis indicate that facilities with larger storage capacity are likely to have a greater number of oil spills, larger volumes of oil spilled, and greater cleanup costs. Similar increases were found at facilities with more tanks and greater annual throughput. Exhibit 1 summarizes the relationships between facility characteristics and spill incident measures that were found to be statistically significant. In addition, EPA's analysis revealed that oil storage facilities in different industry sectors vary in their total storage capacity, number of tanks, and annual throughput volume. Consumption facilities, distribution facilities, production facilities, farms, and institutional facilities, all of which may store oil, do not necessarily conduct their storage operations in the same way.

The results of the analysis also appear to indicate that there are no statistically significant relationships between certain other facility characteristics and spill risk. In particular, EPA did not identify a strong and stable relationship between the type of business conducted at a facility and the number of spills or volume of oil spilled. For example, the difference in volume of oil spilled between farms, distribution, and consumption facilities was not, in general, statistically significant when controlling for other facility characteristics, such as storage capacity, number of tanks, and tank age. The analysis also revealed that the average age of a facility's tanks, the annual number of transfers, and the annual average tank turnover[1] do not appear to be strongly related to oil spills.

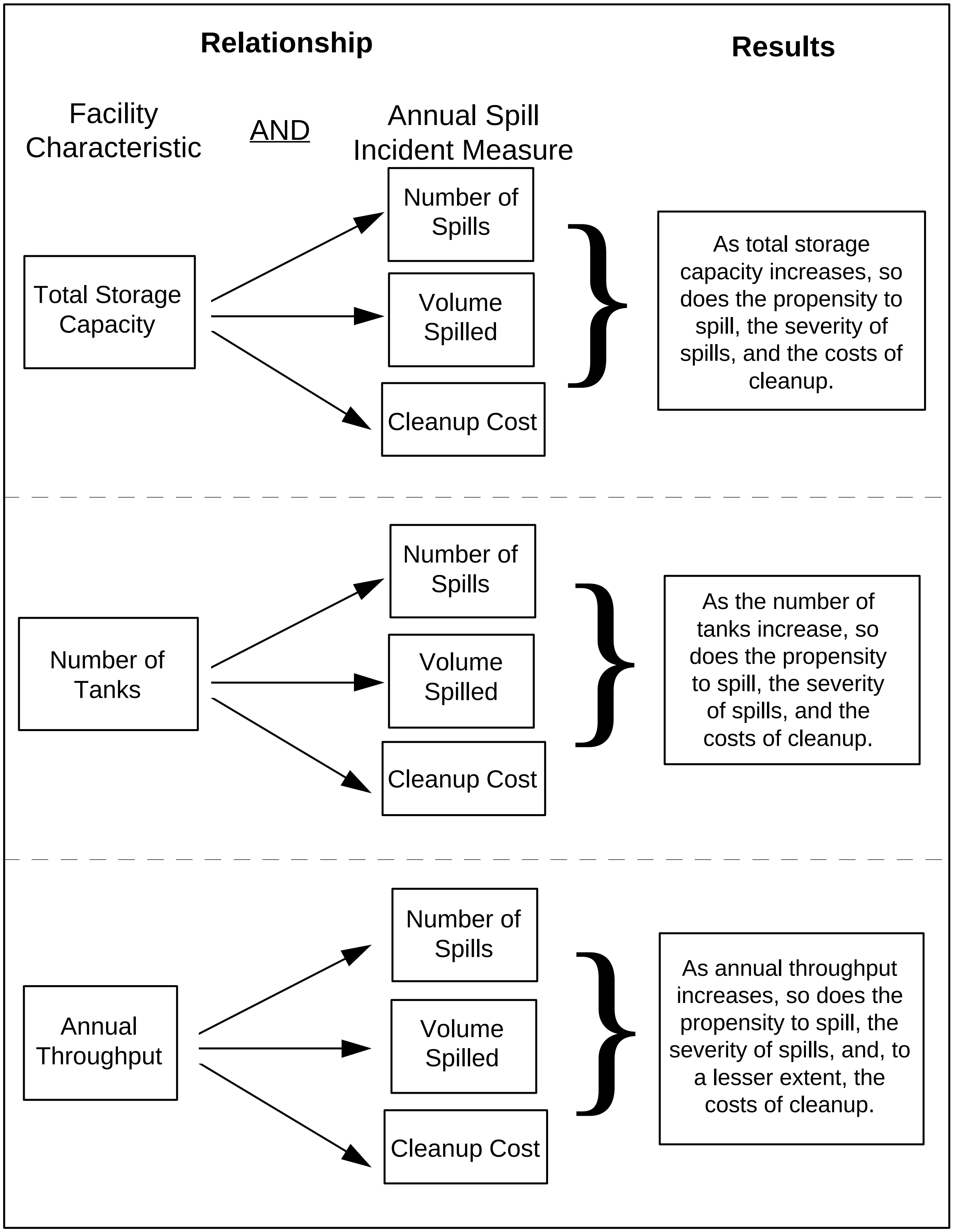
2. APPROACH

EPA conducted statistical analyses to examine possible relationships between certain facility characteristics and spill incidents. For example, EPA wanted to know whether there was a relationship between the number of tanks at a facility and the number of oil spills at that facility. If a relationship were found, EPA wanted to know how significant that relationship was.[2] The facility characteristics that were examined included total storage capacity, number of tanks, annual throughput, facility type, tank age, number of transfers, and average rate of tank turnover. The spill information examined included the annual volume of oil spilled, number of

[1] Tank turnover provides an estimate of the number of times a facility fills its tanks during the year.

[2] Detailed information on the analysis can be found in the appendices to this paper. Specifically, Appendix A provides information on the regression methodologies used to analyze the data and approaches for addressing missing data and other data preparation issues; Appendix B provides the specific numerical results of the regression analyses that show statistically significant relationships; Appendix C presents results of additional regression analyses performed to verify the strength of the analyses; and Appendix D provides information supporting the analysis of industry sectors.

Exhibit 1:
Summary of Significant Relationships



spills, and cleanup costs.[3] For these analyses, EPA conducted both univariate regression analyses on one spill characteristic at a time and multivariate regression analyses to take into account the potential influence of other factors.

3. RESULTS

EPA's regression analysis indicated nine statistically significant relationships. Specifically, the Agency determined that three facility characteristics (total storage capacity, number of tanks, and annual throughput) are each strongly related to three spill incident characteristics (annual volume spilled, number of spills, and cleanup costs). These relationships are discussed in Sections 3.1 through 3.3 below. EPA also identified facility characteristics that do not correlate significantly with spill incidents, which are discussed in Section 3.4. In addition, EPA found certain trends in facility characteristics that correlate to industry types, which are discussed in Section 3.5.

3.1 Facility Storage Capacity

EPA's analysis indicates that the greater the total storage capacity at a facility, the greater the likelihood of having larger spills, more spills, and greater cleanup costs annually. The regression analysis showed that each of these relationships was statistically significant with large R^2 values.[4] For example, the R^2 value of the relationship between the total storage capacity and the number of spills was 0.79 (out of 1), which is considered significant in statistical terms. The R^2 for the relationships between total storage capacity and spill volume and cleanup cost were 0.74 and 0.45, respectively. Exhibit 2 provides a graphical representation of the upward and positive relationship between total storage capacity and the three spill characteristics using four different log values of storage capacity. Specific results from the regression analysis show that a one percent increase in the facility total storage capacity correlates with an estimated 0.5 percent increase in the number of spills annually. Similarly, a one percent increase in storage capacity correlates with a 0.9 percent increase in both volume spilled and cleanup costs annually. The results from EPA's multivariate regression analysis verified that storage capacity is related to the three spill characteristics at statistically significant levels.

In interpreting these results, it is important to keep in mind that the models used in this analysis were meant only to establish whether there is a statistical relationship between the spill incident data and the facility characteristics. For example, the results in Exhibit 2 should not be used to predict or forecast the absolute annual spill volume associated with a given storage

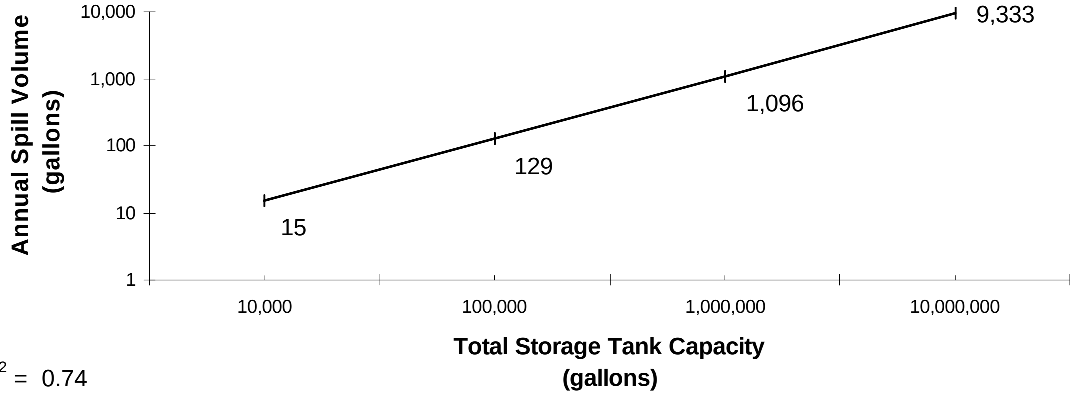
---

[3] Where possible, EPA conducted a comparison of the 1995 SPCC Survey results with the results of the "1994 Survey of API Members' Aboveground Storage Tank (AST) Facilities" (American Petroleum Institute, 1994) to help assess the validity of EPA's results. Except where noted, EPA found general agreement with the results of the two surveys in terms of tank characteristics, spill prevention measures, and sources of spills.

[4] Appendix B of this document provides details on the actual data points and the fitted line, and includes factors supporting the strength of analyses, such as the T-statistic, R^2 values, number of observations, and coefficient estimates. Appendix C presents results of additional analyses performed to further verify the strength of the relationships.

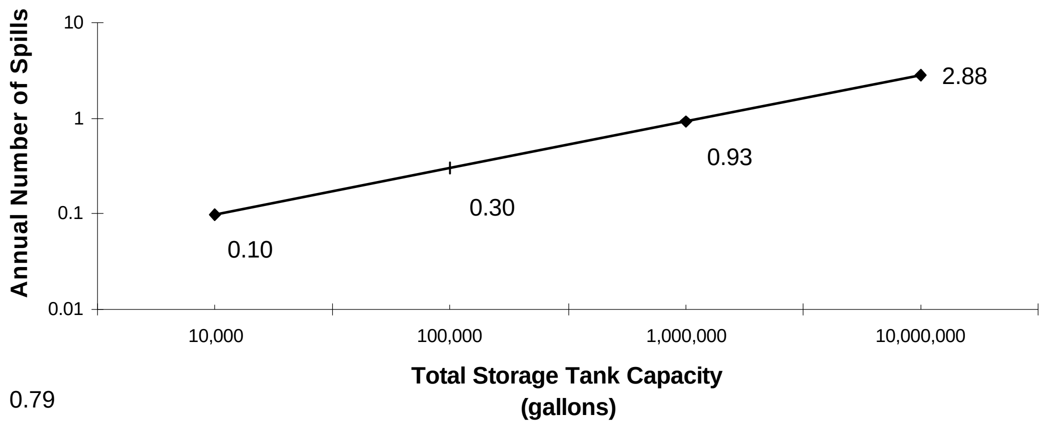
Exhibit 2:

Relationship Between Annual Spill Volume and Total Storage Tank Capacity



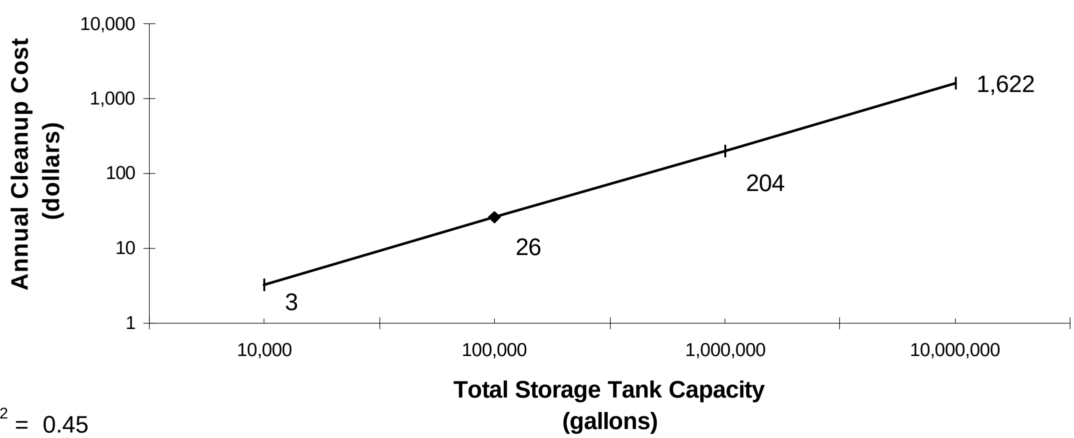
R^2 = 0.74

Relationship Between Annual Number of Spills and Total Storage Tank Capacity



R^2 = 0.79

Relationship Between Annual Cleanup Cost and Total Storage Tank Capacity



R^2 = 0.45

capacity. Rather, the results would support the hypothesis that, based on the survey data, there is a strong positive correlation between storage capacity at a facility and annual spill volume.

3.2 Number of Tanks

EPA's analysis also indicates that as the number of tanks at a facility increases, so does its likelihood of having a larger spill volume, more spills, and greater cleanup costs annually. The regression analyses again showed that each of these relationships was statistically significant; the R^2 values for the relationships between the number of tanks and the spill volume, number of spills, and cleanup costs were 0.50, 0.88, and 0.50, respectively. Exhibit 3 provides a graphical representation of the upward and positive relationship between the number of tanks and the three spill characteristics using three different log values of the number of storage tanks. Specific results from the analysis show that a one percent increase in the number of tanks correlates with an estimated 1.2 percent increase in the spill volume annually. The overall positive relationship is also true for the number of spills and cleanup costs. The results from EPA's multivariate regression analysis verified that number of tanks is related to the three spill characteristics at statistically significant levels.

3.3 Annual Throughput

EPA's analysis also indicates that facilities with greater annual throughput are more likely to have a larger spill volume, more spills, and, to a lesser extent, greater cleanup costs annually. The regression analyses again showed that each of these relationships was statistically significant, although for cleanup costs it was less strong; the R^2 values for the relationships between annual throughput and spill volume, number of spills, and cleanup costs were 0.47, 0.57, and 0.25, respectively. Exhibit 4 provides a graphical representation of the upward and positive relationship between annual throughput and the three spill characteristics using four different log values of throughput. Specific results from the analysis show that a one percent increase in throughput correlates with an estimated 0.7 percent increase in the spill volume annually. Similarly, a one percent increase in throughput correlates with a 0.5 percent increase in the number of spills and 0.3 percent increase in cleanup costs annually. The results from EPA's multivariate regression analysis verified that throughput is related to spill volume at statistically significant levels; however, the relationship to the number of spills and cleanup costs is less significant.

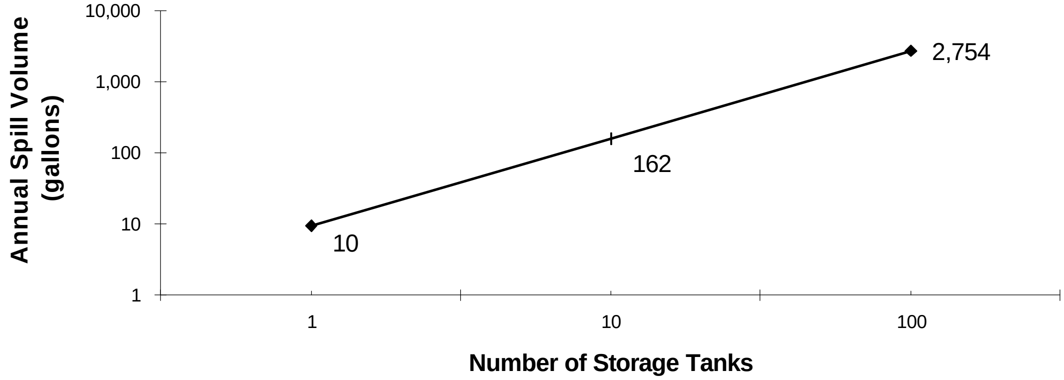
3.4 Other Facility Characteristics

In addition to the three facility characteristics discussed above, EPA investigated other facility characteristics to determine the extent to which they are related to spill incidents. EPA found that facility type, tank age, annual number of transfers, and annual average tank turnover did not have statistically significant relationships to spills.

Facility Type: EPA hypothesized that the number and volume of spills may vary for facilities in different industry sectors. To perform this analysis, EPA organized facilities into the following five groups according to their Standard Industrial Classification (SIC) codes and similarity of operations: farms, oil production facilities, distribution facilities, consumption facilities, and institutional facilities. In general, the analysis indicated that there was no stable

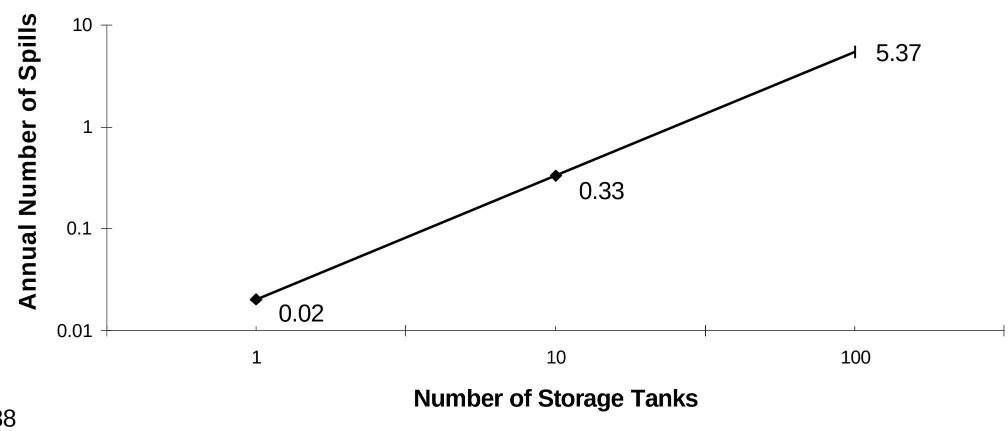
Exhibit 3:

Relationship Between Annual Spill Volume and Number of Storage Tanks



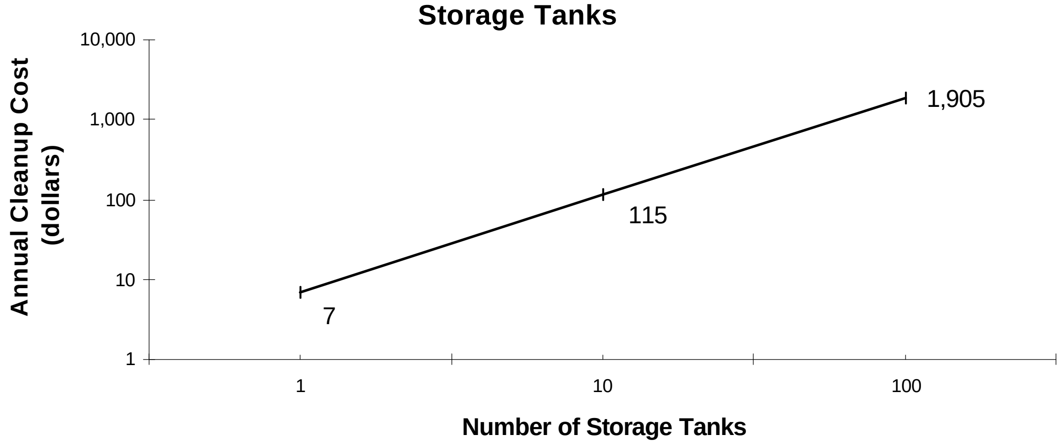
$R^2 = 0.50$

Relationship Between Annual Number of Spills and Number of Storage Tanks



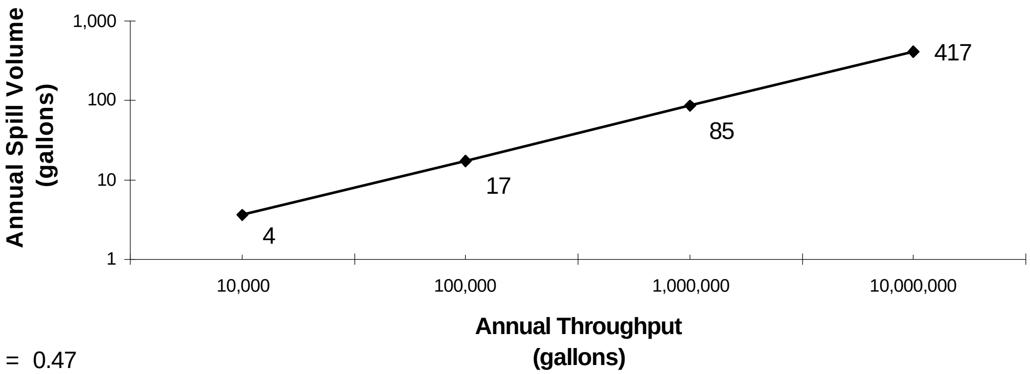
$R^2 = 0.88$

Relationship Between Annual Cleanup Cost and Number of Storage Tanks



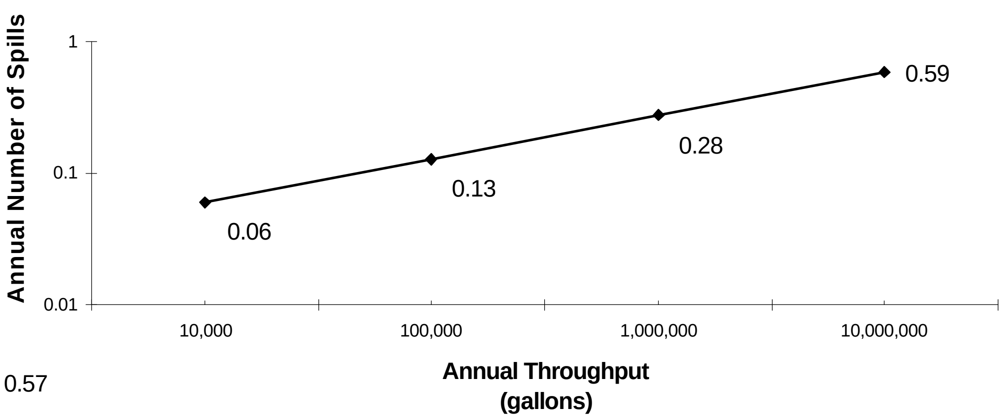
$R^2 = 0.50$

Exhibit 4:
Relationship Between Annual Spill Volume and Annual Throughput



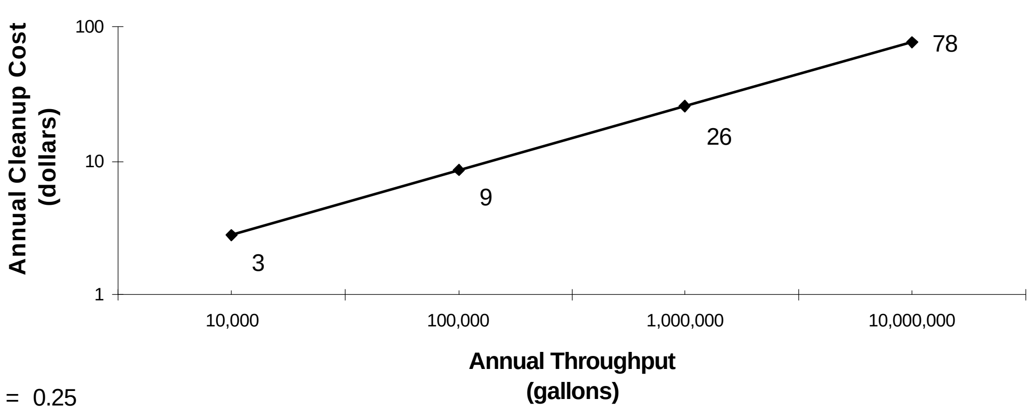
$R^2 = 0.47$

Relationship Between Annual Number of Spills and Annual Throughput



$R^2 = 0.57$

Relationship Between Annual Cleanup Cost and Annual Throughput



$R^2 = 0.25$

and strong relationship between industry groups and number and volume of spills. For example, the difference in volume of oil spilled between farms, distribution, and consumption facilities was not, in general, statistically significant when controlling for other facility characteristics, such as storage capacity, number of tanks, and tank age.

Tank Age: EPA hypothesized that facilities with older tanks may be likely to result in a greater number, volume, and associated cleanup costs of spills because corrosion and related problems become more serious with the passage of time. To analyze this potential relationship for each facility, EPA compared the average age[5] of all of the facility's tanks to the three spill variables.[6] Because many facilities did not provide tank age information, EPA employed several approaches to accommodate missing values for this data element.[7] Regardless of the approach used to address missing data, the results indicate that there is generally a weak relationship between tank age and spill incidents.[8] This relationship diminishes even further when controlling for other, stronger facility characteristics in the regression model.

Annual Number of Transfers: EPA hypothesized that an increase in the annual number of transfers at a facility may increase the frequency and magnitude of spill incidents at that facility. In general, the relationship between the number of transfers and spill incidents at a facility was weak at a statistically significant level, and the importance of this variable in explaining spill incidents diminishes when other facility characteristics are included in the regression model.[9]

Annual Average Tank Turnover: Tank turnover provides an estimate of the number of times a facility fills its tanks during the year, which theoretically reflects a greater level of activity associated with handling or processing oil. EPA hypothesized that spill incidents may increase with tank turnover rate. EPA calculated annual average tank turnover for each facility by dividing annual throughput by storage capacity. The analysis indicated that average tank turnover was only weakly related to spill volume, and not related meaningfully to the other measures of spill incidents. Further analysis of tank turnover and spill incidents for individual SIC code groupings (e.g., farms, distribution) failed to show any meaningful trends.

---

5 Average age was based on weighted-average age using storage capacity.

6 EPA's comparison of the 1995 SPCC Survey results with the 1994 API Survey results indicates that the distribution of tanks by age revealed by EPA's Survey generally is comparable to API's data for facilities in the petroleum refining (SIC 29) and the distribution (i.e., petroleum bulk terminal and stations (SIC 5171) and fuel oil dealers (SIC 5983)) industry sectors, although the 1995 SPCC Survey suggests that there are somewhat newer tanks (15 years or less) in the distribution sector than was identified in the 1994 API Survey.

7 Appendix A discusses the different methods that EPA used.

8 More complex relationships, such as the interactive effect of tank age and tank capacity and spill incidents, were not examined because of the focus on facility-level characteristics and incidents (the investigation of such interactive effects would require the use of data "below," or more detailed than, the level of the facility).

9 Close inspection of the data reveals that a number of respondents are likely to have misinterpreted the question and may have counted the number of times they filled a gas can from the storage tank, for example, as a transfer operation. For example, some small facilities reported a very large number of transfers. This type of measurement error typically biases the coefficient estimate in the regression equation downward and, therefore, may mask the importance of this variable in explaining spill incidents.

3.5 Industry Sector Trends

EPA's analyses revealed that the proportion of facilities that exhibit different characteristics vary by industry sector. In particular, EPA found that facility characteristics, such as the total storage capacity, the number of tanks at a facility, and the annual throughput, vary across the following five industry sector groups: consumption facilities, distribution facilities, production facilities, farms, and institutional facilities.[10] In particular, the analysis indicates that the proportion of facilities with different storage capacities varies significantly between industry sectors. For example, the overwhelming majority (over 80 percent) of facilities in both the farm and institutional industry sectors are small, storing less than 10,000 gallons of oil. In comparison, few (less than 20 percent) in the distribution and production industry sectors are small. Most facilities in the production industry sector store between 10,000 and 50,000 gallons of oil, and most in the distribution industry sector store between 50,000 and 100,000 gallons of oil. Exhibit 5 graphically presents the proportion of facilities in each industry sector according to storage capacity.[11]

In terms of the number of storage tanks present at a facility, EPA's analysis also revealed several trends. In particular, most of the facilities in each industry sector have fewer than five storage tanks. However, facilities in the distribution, consumption, and institutional industry sectors are likely to have more storage tanks than farm or production facilities. For example, 95 percent of the facilities in the farm and production industry sectors have less than ten storage tanks, and none of the facilities in these industries have fifty or more storage tanks. Conversely, there are many facilities in the consumption, distribution, and institutional industry sectors that have more than ten storage tanks, and almost 5 percent of the facilities in these industries have more than fifty storage tanks. Exhibit 6 graphically presents the proportion of facilities in each industry sector according to the number of storage tanks present.

EPA's analysis of annual throughput by industry sector identified several trends. Specifically, of all of the industry sectors analyzed, the farm sector has the greatest percentage of facilities (80 percent) with fewer than 10,000 gallons of annual throughput volume. At the other end of the spectrum, the majority of facilities in the production, distribution, and consumption industry sectors have annual throughput volumes of greater than 100,000 gallons. A majority of facilities in the distribution industry sector (55 percent) have annual throughput volumes of greater than 1 million gallons. In the other industry sectors, such as consumption and institutional, the volumes of annual throughput appear to be more evenly distributed. Exhibit 7 graphically presents the proportion of facilities in each industry sector according to annual throughput.

---

[10] EPA established the five industry sectors according to groupings of Standard Industrial Classification (SIC) codes. Appendix D shows how individual SIC Codes were grouped into the five industry sectors.

[11] EPA's comparison of the 1995 SPCC Survey results with the 1994 API Survey results indicates that the distribution of tanks by storage capacity revealed by the 1995 SPCC Survey generally is comparable to API's data for facilities in the petroleum refining (SIC 29) and the distribution (i.e., petroleum bulk terminal and stations (SIC 5171) and fuel oil dealers (SIC 5983)) industry sectors. The 1995 SPCC Survey suggests that there are somewhat more distribution facilities that have tanks with smaller storage capacities than was identified in the 1994 API Survey.

Exhibit 5:
Proportion of Facilities in Each Industry Sector
According to Storage Capacity

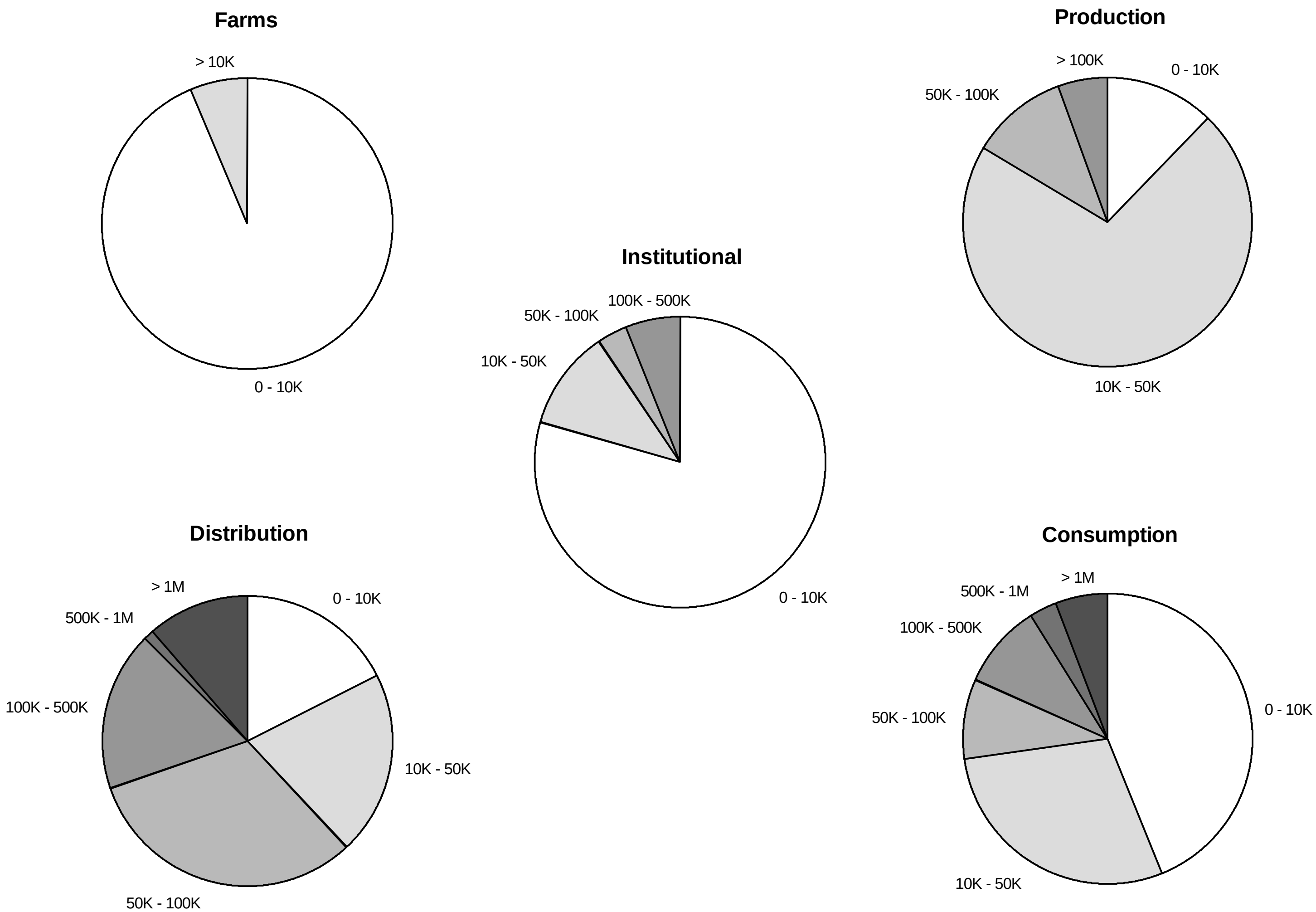


Exhibit 6:
Proportion of Facilities in Each Industry Sector
According to Number of Storage Tanks

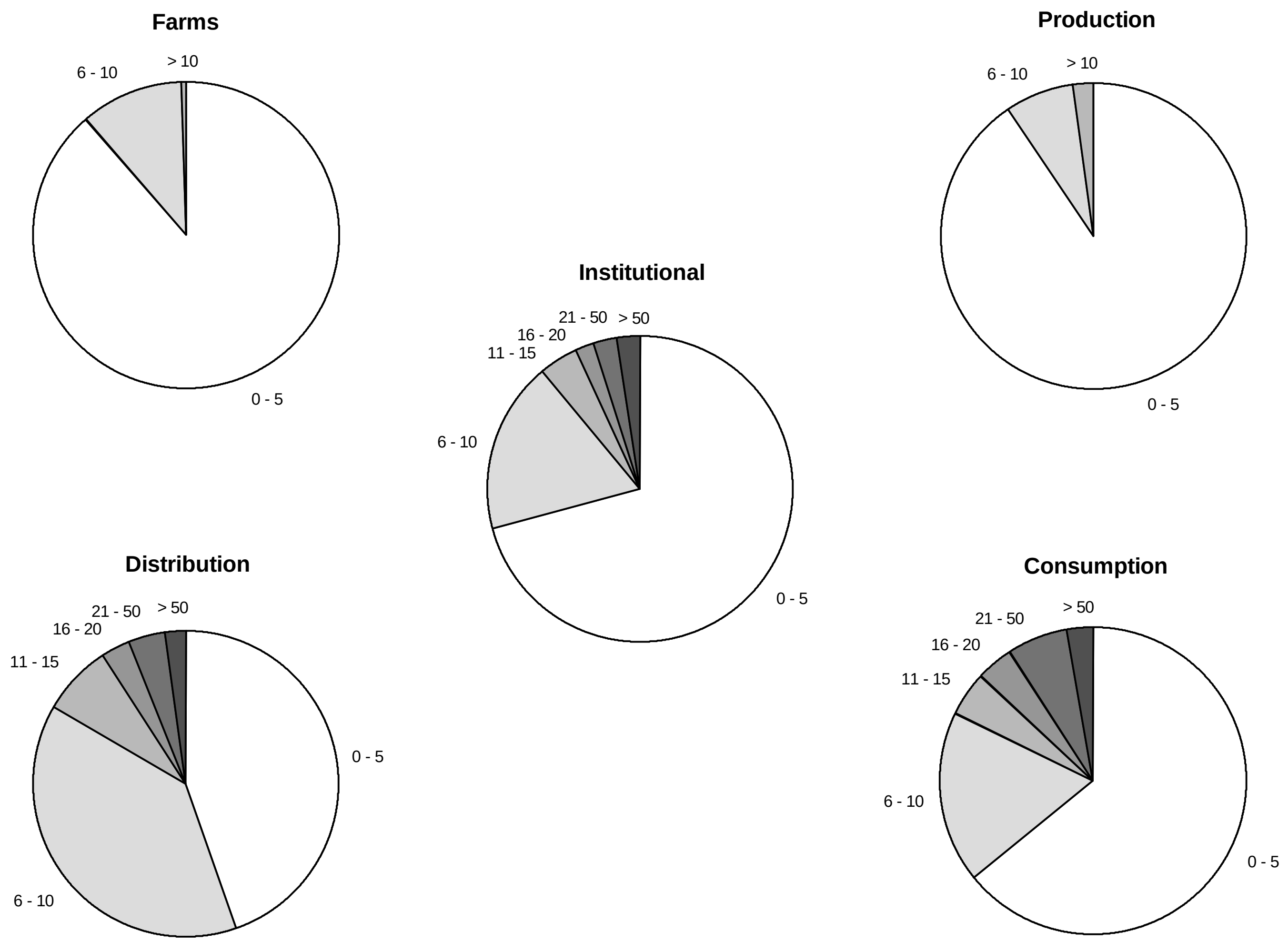
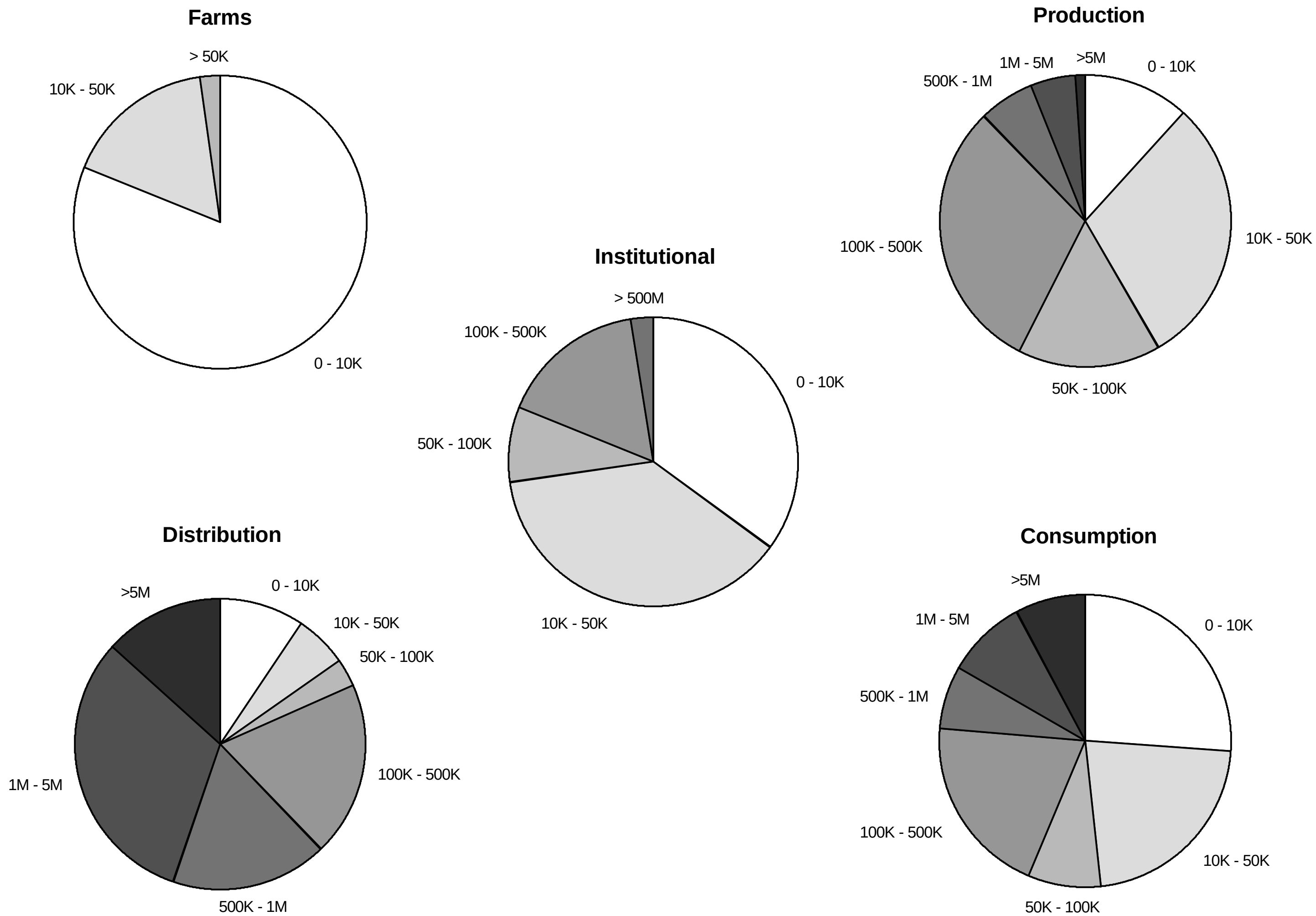


Exhibit 7:
Proportion of Facilities in Each Industry Sector
According to Annual Throughput



APPENDIX A: REGRESSION METHODOLOGY

This appendix provides detailed information on the regression methodologies used to analyze the data, and discusses approaches for addressing missing data and other data preparation issues.

A.1 Regression Models Used in the Analysis

To determine whether there is a relationship between facility characteristics and spill incidents, EPA conducted statistical analyses on the following variables:

Facility Characteristics (Independent Variable)	**Spill Incident Measure (Dependent Variable)**
Total storage capacity	Total annual volume of oil spilled
Number of tanks	Total annual number of spills
Annual throughput	Total annual cleanup expenditures
Facility type (based on SIC code)	
Storage tank age	
Annual number of transfers	
Annual average tank turnover rate	

EPA conducted regression analyses using the Ordinary Least Square (OLS) model to determine the relationships among the facility characteristics and spill incidents. Using OLS, EPA first performed univariate analyses comparing one facility characteristic (independent) variable to one spill incident (dependent) variable. For example, EPA compared the number of tanks to annual volume spilled to determine whether the number of tanks at a facility has an effect on the volume of oil spilled. For each regression model, values for the dependent and independent variables were transformed using log base 10 before the OLS regressions were run; for example, a log base 10 was used for storage capacity rather than actual storage capacity. The purpose of such transformations is to make the relationship between the explanatory variable and the dependent variable linear. Consequently, the parameter estimates presented in this analysis reflect this transformation.

EPA also performed multivariate analyses comparing one spill incident variable to several facility characteristic variables at one time. EPA conducted the multivariate analysis to ensure that single facility variables that show a relationship to spill variables continue to show this relationship even when the potential influence of other facility variables are taken into account.

Additional analyses were performed to address other issues inherent in the data that would have violated certain conditions necessary for using the OLS regression model. For example, in the survey data, the spill parameters (such as total spill volume) cannot assume values less than zero. This results in a limited dependent variable problem. As a result, a truncation in the dispersion in the observations for the spill parameter, for a given value of an independent variable (e.g., storage capacity), occurs when values for the spill parameter approach or become equal to zero. Because this situation violates certain conditions of the OLS regression model, the coefficient estimates in the OLS model will be biased. To account for this, EPA used a Tobit model to determine the extent to which the character of the relationships changed when the problem of truncation was corrected. In general, it was found that the relationships between facility characteristics and spill characteristics identified using the OLS regression models were preserved after the Tobit model was applied, although the magnitude of the coefficient estimates changed. The results of the OLS regressions can, therefore, generally be relied on to establish relationships between facility characteristics and spill characteristics.

EPA also conducted additional analyses to account for issues in using the OLS regression. In general, for the regressions described above, EPA "pooled" or collapsed together all observations for a particular analysis to avoid problems of applying the OLS regression model to a data set that has many zero values for the dependent variable. For example, to account for data points where the total number of spills is zero, EPA averaged the number of spills between facilities with and without spills. The pooling or collapsing of data in this way mitigates the problems of applying the OLS regression model to a data set that has many zero values for the dependent variable (e.g., total annual number of spills) by averaging the number of spills between facilities with and without spills. This approach reduces the variance in the data and may lead to higher R-Squared values compared to the un-pooled data set. Thus, to ensure that this "pooling" did not undermine the strength of the regression analyses, EPA performed the regression models using both pooled and un-pooled data.

A.2 Preparation of the 1995 SPCC Survey Data

Several methods were employed to prepare the survey data prior to performing the analyses. In particular, analysis of the survey information revealed a number of data fields with missing values; for example, the age of the tank often was not provided on the tank table and a number of facilities did not complete the discharge table. Furthermore, some facilities indicated that they have oil storage capacity but failed to complete the tank table. In general, these data issues: (1) limit the number of surveys available for the analysis to only those that provided information on all characteristics of interest; and (2) raise questions about whether reporting bias has occurred. The following methods were used to address these issues in the analyses:

<u>Missing Tank Table</u>. Facilities that completed the survey questionnaire but failed to complete the tank table portion were excluded from the analysis because these records were determined to be unreliable. Of the 2,607 facilities that completed the survey questionnaire, 185 facilities failed to complete the tank table even though they indicated that oil was stored at the facility. The exclusion of these observations from the analysis is not expected to introduce any biases into the results because of the relatively small number involved.

<u>Missing Data Elements Used in the Analysis</u>. When examining multiple factors, missing observations were assigned a value in cases where facilities did not provide certain facility characteristics (e.g., the age of the storage tank, number of transfers, etc.). This approach maintained a high number of observations without compromising the results; otherwise, the number of observations would be limited to those available for the characteristic with the fewest available observations. In addition, EPA determined that certain facilities failed to provide age

data for either all or some of their tanks. EPA developed and applied the following approaches for addressing missing tank age data:

- Eliminating the observations (i.e., surveys) where the data were incomplete;

- Estimating the average age of the facility's tanks using only those tanks for which age data were provided; and

- Assigning a value for tank age. Approaches for assigning age values that were used included: (1) assigning a high estimate (i.e., 75 years) based on the belief that tanks for which age data were not provided are old tanks (and thus tank age is not known by the facility); or (2) assigning the average tank age estimated from those facility's tanks for which this information was provided.

Missing Spill Data Table. A number of facilities completed all portions of the survey questionnaire, but did not positively indicate whether a spill occurred at their facility within the previous year. Because spill underreporting is a serious data issue that could potentially bias the results of the analyses considered in this analysis, this issue was addressed in greater detail. Specifically, an analysis was performed to determine whether the underreporting of spills was a significant problem. A random sample of 300 surveys was generated, which included both facilities that reported and did not report spills. For the group of facilities that did not report any spills, the Emergency Response Notification Systems (ERNS) was searched to determine the extent to which this group of facilities had reportable spills (according to ERNS) and, therefore, underreported spills for purposes of the survey. Out of 85 facilities that reported no spills, only three had at least one report in ERNS (about 3.4 percent). These results suggest that the underreporting of spills on the 1995 SPCC Survey is not a serious problem.

However, the validity of this conclusion is predicated on the assumption that facilities are aware of the National Response Center (NRC) and, furthermore, that they would notify the NRC in the event of a spill (problems raised by the latter point are somewhat mitigated because anyone may contact the NRC about a spill). If small facilities, for example, are assumed to be less aware of the NRC and the Clean Water Act reporting requirements (due to limited resources for example), then these facilities would be less likely to have spill records in ERNS and the results of the comparison described above would be biased downward. Analysis indicates that facilities not reporting any spills are disproportionately smaller than those facilities reporting at least one spill. For example, about 56 percent of all facilities that reported no spills have a storage capacity of less than 10,000 gallons while only 17 percent of all facilities that reported at least one spill have a storage capacity of less than 10,000 gallons.

Although the issue of spill underreporting on the survey cannot be resolved definitively here, in the absence of strong indications from the ERNS cross-matching exercise that the problem of underreporting is highly prevalent, it is reasonable to include all facilities in the analysis including those reporting no discharges.

Trends in Spill Data and Small "Cell" Sizes. Because only one year of data were gathered from facilities, it is possible that for an individual facility the number of spills is not representative, on average, of the annual number of spills (and corresponding spill volume) occurring at the facility due to the episodic nature of spill events and their relatively low frequency of occurrence. For very large samples, this problem is mitigated since some facilities may have experienced a disproportionate number of spills for the survey reporting period

compared to their average over time, while other facilities may have experienced fewer spills that are "typical." Although the available number of surveys is 2,607, which is a satisfactory number to conduct analyses of the overall group, the number of surveys decreases rapidly as the sample is stratified by two or more facility characteristics, such as storage capacity, number of tanks, and SIC code category. As a result, the number of facility surveys in a particular "cell" defined by the three variables (e.g., farm with between 5,000 and 8,000 gallons of total oil storage capacity and three tanks) may be small (e.g., less than ten) and the corresponding number and volume of oil spills for this group of facilities obtained from the 1995 SPCC Survey may diverge from a more typical level (based on averaging over time). The key impact of limited cell sizes is that it is more difficult to establish accurately the extent to which a facility characteristic and spill incident varies. In particular, it is difficult to draw conclusions on the importance of certain factors in explaining spill incidents when they are moderately statistically significant.

APPENDIX B: REGRESSION RESULTS

As previously discussed, EPA identified nine statistically significant relationships. Specifically, EPA determined that each of the three facility characteristic or independent variables (number of tanks, total storage capacity, and annual throughput) have strong relationships with the same three dependent variables (total annual volume spilled, total annual number of spills, and total annual cleanup costs). This appendix provides the detailed results of the analyses for these nine statistically significant relationships.

B.1 Data Points and Predicted Values

Tables providing the actual data points and the points generated by the regression model for each of the nine relationships analyzed are presented in Exhibits B-1 through B-3, as follows:

- **Exhibit B-1** provides the results for the analysis conducted on the total facility storage capacity;
- **Exhibit B-2** provides the results for the analysis conducted on the number of storage tanks located at a facility; and
- **Exhibit B-3** provides the results for the analysis conducted on the total facility throughput.

B.2 Supporting Documentation for Regression Analyses

This section provides the actual results for the nine regression analyses, including factors supporting the strength of the analyses such as the T-statistic, the R^2, the number of observations, and the coefficient estimates for each of the analysis, which indicates the degree to which an increase in each facility characteristic affects the particular spill incident variable being analyzed.

Number of Tanks

A. The log-log regression model that relates the total annual spill volume to the number of storage tanks at the facility is summarized below:

Dependent variable: The total annual spill volume in gallons.

Independent variable: The number of storage tanks.

Intercept:	0.98
Coefficient estimate:	1.23
T-statistic:	3.49
R2:	0.50
Number of observations:	14

EXHIBIT B-1:
ACTUAL AND PREDICTED SPILL VOLUME BY STORAGE CAPACITY

TOTAL STORAGE TANK CAPACITY (gallons)	ACTUAL STORAGE TANK CAPACITY (gallons)	ACTUAL SPILL VOLUME (gallons)	PREDICTED SPILL VOLUME (gallons)
0-500	283	1.29	0.55
500-1K[1]	928	1.78	1.66
1K-1.5K	1,367	1.57	2.38
1.5K-2K	1,812	0.59	3.09
2K-2.5K	2,293	0.85	3.85
2.5K-3K	2,814	0.09	4.66
3K-4K	3,573	6.03	5.81
4K-5K	4,606	0.63	7.36
5K-6K	5,612	34.12	8.85
6K-8K	7,253	4.73	11.23
8K-10K	9,077	43.12	13.84
10K-12K	11,134	3.19	16.73
12K-14K	12,859	211.54	19.13
14K-17K	16,478	61.09	24.08
17K-20K	18,036	6.94	26.20
20K-25K	22,293	624.45	31.90
25K-30K	26,434	61.90	37.38
30K-35K	33,008	200.43	45.95
35K-40K	37,685	323.39	51.97
40K-50K	44,450	38.14	60.60
50K-60K	54,665	540.18	73.45
60K-70K	64,651	46.26	85.85
70K-80K	74,845	156.48	98.37
80K-100K	88,662	783.61	115.15
100K-150K	122,130	149.21	155.09
150K-200K	172,947	320.24	214.32
200K-400K	276,064	351.83	331.05
400K-1M[2]	633,615	223.38	716.75
1M-5M	2,240,652	2,236.21	2,319.48
5M-10M	6,534,725	2,880.73	6,274.76
Greater than 10M	207,676,305	38,220.58	156,408.25

[1] The symbol K represents 1,000.
[2] The symbol M represents 1,000,000.

EXHIBIT B-1:
ACTUAL AND PREDICTED NUMBER OF SPILLS BY STORAGE CAPACITY

TOTAL STORAGE TANK CAPACITY (gallons)	ACTUAL STORAGE TANK CAPACITY (gallons)	ACTUAL NUMBER OF SPILLS	PREDICTED NUMBER OF SPILLS
0-500	283	0.06	0.02
500-1K[1]	928	0.02	0.03
1K-1.5K	1,367	0.04	0.04
1.5K-2K	1,812	0.02	0.04
2K-2.5K	2,293	0.03	0.05
2.5K-3K	2,814	0.03	0.05
3K-4K	3,573	0.06	0.06
4K-5K	4,606	0.05	0.07
5K-6K	5,612	0.18	0.08
6K-8K	7,253	0.07	0.09
8K-10K	9,077	0.09	0.10
10K-12K	11,134	0.10	0.11
12K-14K	12,859	0.12	0.12
14K-17K	16,478	0.24	0.13
17K-20K	18,036	0.12	0.14
20K-25K	22,293	0.29	0.15
25K-30K	26,434	0.13	0.17
30K-35K	33,008	0.04	0.18
35K-40K	37,685	0.11	0.20
40K-50K	44,450	0.16	0.21
50K-60K	54,665	0.25	0.24
60K-70K	64,651	0.89	0.26
70K-80K	74,845	0.20	0.28
80K-100K	88,662	0.15	0.30
100K-150K	122,130	0.76	0.35
150K-200K	172,947	0.80	0.42
200K-400K	276,064	0.88	0.53
400K-1M[2]	633,615	1.59	0.79
1M-5M	2,240,652	5.34	1.48
5M-10M	6,534,725	1.93	2.51
Greater than 10M	207,676,305	4.38	13.85

[1] The symbol K represents 1,000.
[2] The symbol M represents 1,000,000.

EXHIBIT B-1:
ACTUAL AND PREDICTED CLEANUP COST BY STORAGE CAPACITY

TOTAL STORAGE TANK CAPACITY (gallons)	ACTUAL STORAGE TANK CAPACITY (gallons)	ACTUAL CLEANUP COST (dollars)	PREDICTED CLEANUP COST (dollars)
0-500	283	0.03	0.13
500-1K[1]	928	0.04	0.37
1K-1.5K	1,367	0.11	0.52
1.5K-2K	1,812	4.54	0.67
2K-2.5K	2,293	0.00	0.83
2.5K-3K	2,814	0.00	1.00
3K-4K	3,573	189.12	1.23
4K-5K	4,606	0.05	1.55
5K-6K	5,612	81.55	1.85
6K-8K	7,253	0.42	2.33
8K-10K	9,077	9.14	2.84
10K-12K	11,134	0.05	3.41
12K-14K	12,859	219.44	3.88
14K-17K	16,478	24.59	4.85
17K-20K	18,036	3.06	5.26
20K-25K	22,293	24.78	6.36
25K-30K	26,434	51.60	7.41
30K-35K	33,008	63.91	9.04
35K-40K	37,685	5.86	10.18
40K-50K	44,450	1.13	11.80
50K-60K	54,665	323.56	14.20
60K-70K	64,651	0.11	16.50
70K-80K	74,845	0.18	18.82
80K-100K	88,662	0.11	21.90
100K-150K	122,130	250.64	29.18
150K-200K	172,947	160.16	39.84
200K-400K	276,064	112.38	60.57
400K-1M[2]	633,615	269.25	127.48
1M-5M	2,240,652	2,897.17	395.15
5M-10M	6,534,725	5,802.40	1,030.69
Greater than 10M	207,676,305	3,323.88	22,835.14

[1] The symbol K represents 1,000.
[2] The symbol M represents 1,000,000.

EXHIBIT B-2:
ACTUAL AND PREDICTED SPILL VOLUME BY NUMBER OF TANKS

TOTAL NUMBER OF STORAGE TANKS	ACTUAL NUMBER OF STORAGE TANKS	ACTUAL SPILL VOLUME (gallons)	PREDICTED SPILL VOLUME (gallons)
1	1	82.04	9.58
2	2	169.92	22.53
3	3	31.56	37.16
4	4	62.69	53.00
5	5	18.25	69.79
6	6	18.78	87.40
7	7	261.37	105.71
8	8	14.10	124.64
9	9	72.98	144.14
10	10	19.16	164.15
11-15	13	203.97	216.72
15-20	18	592.00	330.82
21-50	33	798.42	708.05
Greater than 50	141	37,707.19	4,314.35

EXHIBIT B-2:
ACTUAL AND PREDICTED NUMBER OF SPILLS BY NUMBER OF TANKS

TOTAL NUMBER OF STORAGE TANKS	ACTUAL NUMBER OF STORAGE TANKS	ACTUAL NUMBER OF SPILLS	PREDICTED NUMBER OF SPILLS
1	1	0.04	0.02
2	2	0.07	0.05
3	3	0.11	0.08
4	4	0.09	0.11
5	5	0.18	0.14
6	6	0.08	0.18
7	7	0.23	0.21
8	8	0.09	0.25
9	9	0.18	0.29
10	10	0.18	0.33
11-15	13	0.76	0.43
15-20	18	0.79	0.65
21-50	33	2.25	1.37
Greater than 50	141	10.69	8.01

EXHIBIT B-2:
ACTUAL AND PREDICTED CLEANUP COST BY NUMBER OF TANKS

TOTAL NUMBER OF STORAGE TANKS	ACTUAL NUMBER OF STORAGE TANKS	ACTUAL CLEANUP COST (dollars)	PREDICTED CLEANUP COST (dollars)
1	1	12.75	6.96
2	2	28.11	16.18
3	3	52.75	26.51
4	4	4.00	37.63
5	5	33.46	49.38
6	6	48.79	61.65
7	7	154.19	74.38
8	8	194.87	87.51
9	9	1,224.76	101.00
10	10	6.11	114.83
11-15	13	190.53	151.05
15-20	18	32.79	229.28
21-50	33	1,805.58	485.82
Greater than 50	141	4,008.58	2,890.25

EXHIBIT B-3:
ACTUAL AND PREDICTED SPILL VOLUME BY THROUGHPUT

TOTAL THROUGHPUT (gallons)	ACTUAL THROUGHPUT (gallons)	ACTUAL SPILL VOLUME (gallons)	PREDICTED SPILL VOLUME (gallons)
0-100	4	9.23	0.01
100-1K[1]	624	1.21	0.52
1K-2K	1,681	8.46	1.03
2K-3K	2,679	90.45	1.42
3K-4K	3,766	0.01	1.79
4K-5K	4,787	0.01	2.11
5K-6K	5,745	1.15	2.39
6K-7K	6,740	0.20	2.67
7K-8K	7,786	8.13	2.95
8K-9K	8,718	0.47	3.19
9K-10K	9,915	0.03	3.49
10K-12K	11,540	2.82	3.87
12K-15K	13,941	0.07	4.41
15K-20K	17,578	3.41	5.17
20K-25K	23,046	1.71	6.24
25K-30K	28,694	0.00	7.25
30K-40K	35,082	37.35	8.33
40K-50K	45,820	180.13	10.01
50K-100K	73,779	100.56	13.90
100K-200K	125,818	34.95	20.08
200K-300K	176,602	12.27	25.36
300K-400K	254,261	557.67	32.60
400K-500K	403,829	401.25	44.83
500K-750K	611,455	13.33	59.67
750K-1M[2]	884,226	121.89	76.93
1M-2M	1,482,710	78.71	109.84
2M-3M	2,525,511	517.53	158.53
3M-5M	3,966,784	367.49	216.37
5M-10M	7,428,815	348.20	333.37
10M-100M	41,509,536	3,976.74	1,090.73
Greater than 100M	1,172,996,009	40,215.55	10,901.04

[1] The symbol K represents 1,000.
[2] The symbol M represents 1,000,000.

EXHIBIT B-3:
ACTUAL AND PREDICTED SPILL VOLUME BY THROUGHPUT

TOTAL THROUGHPUT (gallons)	ACTUAL THROUGHPUT (gallons)	ACTUAL NUMBER OF SPILLS	PREDICTED NUMBER OF SPILLS
0-100	4	0.05	0.00
100-1K[1]	624	0.03	0.02
1K-2K	1,681	0.55	0.03
2K-3K	2,679	0.03	0.04
3K-4K	3,766	0.05	0.05
4K-5K	4,787	0.01	0.05
5K-6K	5,745	0.01	0.05
6K-7K	6,740	0.04	0.05
7K-8K	7,786	0.01	0.06
8K-9K	8,718	0.07	0.06
9K-10K	9,915	0.02	0.06
10K-12K	11,540	0.18	0.07
12K-15K	13,941	0.03	0.07
15K-20K	17,578	0.04	0.08
20K-25K	23,046	0.02	0.08
25K-30K	28,694	0.04	0.09
30K-40K	35,082	0.11	0.10
40K-50K	45,820	0.22	0.10
50K-100K	73,779	0.13	0.12
100K-200K	125,818	0.19	0.15
200K-300K	176,602	0.19	0.16
300K-400K	254,261	0.51	0.18
400K-500K	403,829	0.40	0.22
500K-750K	611,455	0.14	0.25
750K-1M[2]	884,226	0.21	0.28
1M-2M	1,482,710	0.21	0.33
2M-3M	2,525,511	0.21	0.40
3M-5M	3,966,784	0.84	0.46
5M-10M	7,428,815	0.77	0.57
10M-100M	41,509,536	5.33	1.01
Greater than 100M	1,172,996,009	4.91	3.08

[1] The symbol K represents 1,000.
[2] The symbol M represents 1,000,000.

EXHIBIT B-3:
ACTUAL AND PREDICTED CLEANUP COST BY THROUGHPUT

TOTAL THROUGHPUT (gallons)	ACTUAL THROUGHPUT (gallons)	ACTUAL CLEANUP COST (dollars)	PREDICTED CLEANUP COST (dollars)
0-100	4	353.09	0.06
100-1K[1]	624	0.04	0.76
1K-2K	1,681	0.06	1.22
2K-3K	2,679	86.96	1.53
3K-4K	3,766	0.10	1.80
4K-5K	4,787	0.00	2.02
5K-6K	5,745	0.00	2.21
6K-7K	6,740	0.00	2.38
7K-8K	7,786	281.28	2.55
8K-9K	8,718	0.04	2.70
9K-10K	9,915	0.02	2.87
10K-12K	11,540	2.10	3.08
12K-15K	13,941	0.00	3.38
15K-20K	17,578	0.05	3.78
20K-25K	23,046	11.96	4.30
25K-30K	28,694	0.00	4.78
30K-40K	35,082	23.22	5.27
40K-50K	45,820	8.39	5.99
50K-100K	73,779	56.23	7.53
100K-200K	125,818	44.90	9.74
200K-300K	176,602	11.38	11.46
300K-400K	254,261	1.14	13.66
400K-500K	403,829	97.10	17.07
500K-750K	611,455	0.06	20.84
750K-1M[2]	884,226	149.36	24.88
1M-2M	1,482,710	65.66	31.91
2M-3M	2,525,511	353.70	41.23
3M-5M	3,966,784	52.91	51.24
5M-10M	7,428,815	328.27	69.29
10M-100M	41,509,536	2,949.85	158.57
Greater than 100M	1,172,996,009	3,671.55	791.53

[1] The symbol K represents 1,000.
[2] The symbol M represents 1,000,000.

Number of Tanks (cont'd)

The coefficient estimate indicates that a one percent increase in the number of tanks at the facility results in a 1.23 percent increase in the total annual spill volume.

B. The log-log regression model that relates the total annual number of reported spills to the number of storage tanks at the facility is summarized below:

Dependent variable: The total annual number of reported spills.

Independent variable: The number of storage tanks.

Intercept:	-1.69
Coefficient estimate:	1.21
T-statistic:	9.49
R2:	0.88
Number of observations:	14

The coefficient estimate indicates that a one percent increase in the number of tanks at the facility results in a 1.21 percent increase in the total annual number of spills.

C. The log-log regression model that relates the total annual cleanup cost to the number of storage tanks at the facility is summarized below:

Dependent variable: The total annual cleanup cost.

Independent variable: The number of storage tanks.

Intercept:	0.84
Coefficient estimate:	1.22
T-statistic:	3.48
R2:	0.50
Number of observations:	14

The coefficient estimate indicates that a one percent increase in the number of tanks at the facility results in a 1.22 percent increase in the total annual cleanup cost.

Total Storage Capacity

A. The log-log regression model that relates the total annual spill volume to the total storage capacity of the facility is summarized below:

Dependent variable: The total annual spill volume in gallons.

Independent variable: The total storage capacity in gallons.

Intercept:	-2.54
Coefficient estimate:	0.93
T-statistic:	9.04
R2:	0.74
Number of observations:	31

The coefficient estimate indicates that a one percent increase in total storage capacity results in a 0.93 percent increase in the total annual spill volume.

B. The log-log regression model that relates the total annual number of reported spills to the total storage capacity of the facility is summarized below:

Dependent variable: The total annual number of reported spills.

Independent variable: The total storage capacity in gallons.

Intercept:	-2.97
Coefficient estimate:	0.49
T-statistic:	10.39
R2:	0.79
Number of observations:	31

The coefficient estimate indicates that a one percent increase in total storage capacity results in a 0.49 percent increase in the total annual number of spills.

C. The log-log regression model that relates the total annual cleanup cost to the total storage capacity of the facility is summarized below:

Dependent variable: The total annual cleanup cost.

Independent variable: The total storage capacity in gallons.

Intercept:	-3.09
Coefficient estimate:	0.90
T-statistic:	4.88
R2:	0.45
Number of observations:	31

The coefficient estimate indicates that a one percent increase in total storage capacity results in a 0.90 percent increase in the total annual cleanup cost.

Annual Throughput

A. The log-log regression model that relates the total annual spill volume to the annual throughput of the facility is summarized below:

Dependent variable: The total annual spill volume in gallons.

Independent variable: The annual throughput in gallons.

Intercept:	-2.21
Coefficient estimate:	0.69
T-statistic:	5.12
R2:	0.47
Number of observations:	31

The coefficient estimate indicates that a one percent increase in the annual throughput results in a 0.69 percent increase in the total annual spill volume.

B. The log-log regression model that relates the total annual number of reported spills to the annual throughput of the facility is summarized below:

Dependent variable: The total annual number of reported spills.

Independent variable: The annual throughput in gallons.

Intercept:	-2.54
Coefficient estimate:	0.33
T-statistic:	6.22
R2:	0.57
Number of observations:	31

The coefficient estimate indicates that a one percent increase in the annual throughput results in a 0.33 percent increase in the total annual number of spills.

C. The log-log regression model that relates the total annual cleanup cost to the annual throughput of the facility is summarized below:

Dependent variable: The total annual cleanup cost.

Independent variable: The annual throughput in gallons.

Intercept:	-1.47
Coefficient estimate:	0.48
T-statistic:	3.12
R2:	0.25
Number of observations:	31

The coefficient estimate indicates that a one percent increase in the annual throughput results in a 0.25 percent increase in the total annual cleanup cost.

APPENDIX C: ADDITIONAL ANALYSES

This appendix presents results of additional regression analyses performed to verify the strength of the analyses. Section C.1 provides the results of the regression analyses performed on "un-pooled" data. Section C.2 provides the results of the multivariate regression analyses.

C.1 Results of Additional Regression Analyses Performed on "Un-Pooled" Data

As previously discussed, because the regression results were developed using pooled data, the variance in the risk variable for a given value of the independent variable is reduced compared to the un-pooled data set. This transformation leads to higher R-Squared values compared to the un-pooled data set. To further verify the strength of the relationships presented above, the same regression results were generated using the un-pooled data. The key parameters associated with the regression equation are presented in the tables below. As expected, the R-Squared values are generally lower for the un-pooled data. However, all of the relationships between facility characteristics and spill risk continue to be highly statistically significant when the un-pooled data are used.

<u>Number of Tanks</u>

A. The un-pooled log-log regression model that relates the total annual spill volume to the number of storage tanks is summarized below:

Dependent variable: The total annual spill volume in gallons.

Independent variable: The number of storage tanks.

Intercept:	-0.17
Coefficient estimate:	0.69
T-statistic:	18.29
R2:	0.12
Number of observations:	2350

B. The un-pooled log-log regression model that relates the total annual number of reported spills to the number of storage tanks is summarized below:

Dependent variable: The total annual number of reported spills.

Independent variable: The number of storage tanks.

Intercept:	-0.04
Coefficient estimate:	0.12
T-statistic:	18.53
R2:	0.13
Number of observations:	2350

C. The un-pooled log-log regression model that relates the total annual cleanup cost to the number of storage tanks is summarized below:

Dependent variable: The total annual cleanup cost.

Independent variable: The number of storage tanks.

Intercept:	-0.08
Coefficient estimate:	0.30
T-statistic:	11.45
R2:	0.05
Number of observations:	2350

Total Storage Capacity

A. The un-pooled log-log regression model that relates the total annual spill volume to the total storage capacity of the facility is summarized below:

Dependent variable: The total annual spill volume in gallons.

Independent variable: The total storage capacity in gallons.

Intercept:	-1.15
Coefficient estimate:	0.33
T-statistic:	21.80
R2:	0.17
Number of observations:	2350

B. The un-pooled log-log regression model that relates the total annual number of reported spills to the total storage capacity of the facility is summarized below:

Dependent variable: The total annual number of reported spills.

Independent variable: The total storage capacity in gallons.

Intercept:	-0.17
Coefficient estimate:	0.05
T-statistic:	16.83
R2:	0.11
Number of observations:	2350

C. The un-pooled log-log regression model that relates the total annual cleanup cost to the total storage capacity of the facility is summarized below:

Dependent variable: The total annual cleanup cost.

Independent variable: The total storage capacity in gallons.

Intercept:	-0.46
Coefficient estimate:	0.13
T-statistic:	12.33
R2:	0.06
Number of observations:	2350

Annual Throughput

A. The un-pooled log-log regression model that relates the total annual spill volume to the annual throughput of the facility is summarized below:

Dependent variable: The total annual spill volume in gallons.

Independent variable: The annual throughput in gallons.

Intercept:	-0.39
Coefficient estimate:	0.14
T-statistic:	15.42
R2:	0.09
Number of observations:	2350

B. The un-pooled log-log regression model that relates the total annual number of reported spills to the annual throughput of the facility is summarized below:

Dependent variable: The total annual number of reported spills.

Independent variable: The annual throughput in gallons.

Intercept:	-0.05
Coefficient estimate:	0.02
T-statistic:	11.24
R2:	0.05
Number of observations:	2350

C. The un-pooled log-log regression model that relates the total annual cleanup cost to the annual throughput of the facility is summarized below:

Dependent variable: The total annual number of reported spills.

Independent variable: The annual throughput in gallons.

Intercept:	-0.14
Coefficient estimate:	0.05
T-statistic:	8.52
R2:	0.03
Number of observations:	2350

C.2 Results of Multivariate Regression Analyses

Because storage capacity, number of tanks, and throughput were identified as important individual factors in explaining the total annual spill volume, number of spills, and cleanup costs, these factors were used together in a multivariate regression model. This step is necessary to ensure that these three variables continue to be statistically significant variables when taking into account the degree of multicollinearity and to correct for Left Out Variable Bias, which can lead to an overstatement of the importance of the variable in explaining the variation in the dependent variable. The results of these analyses are presented in the tables below. In general, storage capacity and number of tanks were found to be (highly) statistically significantly related to all three measures of spill risk (i.e., total number, volume, and cleanup costs of oil spills). However, annual throughput was only related to total annual spill volume at a statistically significant level; throughput was only moderately related to total annual number of spills and total annual cleanup cost expenditures (i.e., this characteristic is significant only at the 85 percent confidence level), as shown in the regression equations.

The un-pooled log-log regression model that relates the total annual number of reported spills to the number of storage tanks, total storage capacity, and annual throughput is summarized below:

Dependent variable:		The total annual number of reported spills.
Independent variable	A:	The number of storage tanks.
Independent variable	B:	The total storage capacity in gallons.
Independent variable	C:	The annual throughput in gallons.
Intercept:		-0.14
Coefficient estimate	A:	0.09
Coefficient estimate	B:	0.03
Coefficient estimate	C:	0.003
T-statistic	A:	11.49
T-statistic	B:	7.12
T-statistic	C:	1.43
R2:		0.16
Number of observations:		2350

The un-pooled log-log regression model that relates the total annual spill volume to the number of storage tanks, total storage capacity, and annual throughput is summarized below:

Dependent variable:		The total annual spill volume in gallons.
Independent variable	A:	The number of storage tanks.
Independent variable	B:	The total storage capacity in gallons.
Independent variable	C:	The annual throughput in gallons.
Intercept:		-1.03
Coefficient estimate	A:	0.37
Coefficient estimate	B:	0.22

Coefficient estimate	C:	0.04
T-statistic	A:	8.74
T-statistic	B:	10.83
T-statistic	C:	3.72
R2:		0.20
Number of observations:		2350

The un-pooled log-log regression model that relates the total annual cleanup cost to the number of storage tanks, total storage capacity, and annual throughput is summarized below:

Dependent variable:		The total annual cleanup cost.
Independent variable	A:	The number of storage tanks.
Independent variable	B:	The total storage capacity in gallons.
Independent variable	C:	The annual throughput in gallons.
Intercept:		-0.40
Coefficient estimate	A:	0.18
Coefficient estimate	B:	0.09
Coefficient estimate	C:	0.01
T-statistic	A:	5.97
T-statistic	B:	5.91
T-statistic	C:	1.43
R2:		0.08
Number of observations:		2350

APPENDIX D:
SIC CODE AND DESCRIPTION BY INDUSTRY SECTOR

Sector	SIC Code	Description
Distribution	2911	Petroleum refining
	4612	Crude petroleum pipelines
	4911	Electric services
	5171	Petroleum bulk stations & terminals
	5541	Gasoline service stations
	5983	Fuel oil dealers
	7513	Truck rental and leasing, no drivers
Production	131	Crude Petroleum And Natural Gas
Farms	01	Agricultural Production--Crops
Institutional	8062	General medical & surgical hospitals
	8063	Psychiatric hospitals
	8211	Elementary and secondary schools
	8221	Colleges and universities
Consumption	1221	Bituminous coal and lignite--surface
	1241	Coal mining services
	1411	Dimension stone
	1422	Crushed and broken limestone
	1429	Crushed and broken stone, nec
	1442	Construction sand and gravel
	1446	Industrial sand
	1455	Kaolin and ball clay
	1611	Highway and street construction
	1623	Water, sewer, and utility lines
	1629	Heavy construction, nec
	2011	Meat packing plants
	2013	Sausages and other prepared meats
	2015	Poultry slaughtering and processing
	2022	Cheese, natural and processed
	2023	Dry, condensed, evaporated products
	2026	Fluid milk
	2035	Pickles, sauces, and salad dressings
	2037	Frozen fruits and vegetables
	2038	Frozen specialties, nec
	2041	Flour and other grain mill products
	2046	Wet corn milling
	2048	Prepared feeds, nec
	2051	Bread, cake, and related products
	2052	Cookies and crackers
	2063	Beet sugar
	2064	Candy & other confectionery products
	2066	Chocolate & cocoa products
	2075	Soybean oil mills
	2077	Animal and marine fats and oils
	2079	Edible fats and oils, nec
	2085	Distilled and blended liquors
	2086	Bottled and canned soft drinks
	2091	Canned and cured fish and seafoods
	2092	Fresh or frozen prepared fish
	2096	Potato chips and similar snacks
	2099	Food preparations, nec
	2221	Broadwoven fabric mills, manmade
	2241	Narrow fabric mills
	2269	Finishing plants, nec
	2273	Carpets and rugs
	2295	Coated fabrics, not rubberized
	2329	Men's & boys' clothing, nec
	2411	Logging
	2421	Sawmills & planing mills, general
	2431	Millwork
	2448	Wood pallets and skids
	2491	Wood preserving
	2493	Reconstituted wood products
	2511	Wood household furniture
	2512	Upholstered household furniture
	2531	Public building & related furniture
	2611	Pulp mills
	2621	Paper mills
	2631	Paperboard mills
	2653	Corrugated & solid fiber boxes
	2679	Converted paper products, nec
	2711	Newspapers
	2731	Book publishing
	2732	Book printing
	2752	Commercial printing, lithographic
	2759	Commercial printing, nec
	2812	Alkalies and chlorine
	2819	Industrial inorganic chemicals, nec
	2821	Plastics materials and resins
	2823	Cellulosic manmade fibers
	2833	Medicinals and botanicals
	2834	Pharmaceutical preparations
	2841	Soap and other detergents
	2842	Polishes and sanitation goods
	2851	Paints and allied products
	2869	Industrial organic chemicals, nec
	2873	Nitrogenous fertilizers
	2874	Phosphatic fertilizers
	2875	Fertilizers, mixing only
	2879	Agricultural chemicals, nec
	2891	Adhesives and sealants
	2893	Printing ink
	2899	Chemical preparations, nec
	2951	Asphalt paving mixtures and blocks
	2952	Asphalt felts and coatings
	2992	Lubricating oils and greases
	3011	Tires and inner tubes
	3086	Plastics foam products
	3089	Plastics products, nec
	3131	Footwear cut stock
	3241	Cement, hydraulic
	3272	Concrete products, nec
	3273	Ready-mixed concrete
	3281	Cut stone and stone products
	3295	Minerals, ground or treated
	3312	Blast furnaces and steel mills
	3315	Steel wire and related products
	3325	Steel foundries, nec
	3353	Aluminum sheet, plate, and foil
	3354	Aluminum extruded products
	3357	Nonferrous wiredrawing & insulating
	3369	Nonferrous foundries, nec
	3411	Metal cans
	3412	Metal barrels, drums, and pails
	3423	Hand and edge tools, nec

3429	Hardware, nec
3443	Fabricated plate work (boiler shops)
3444	Sheet metalwork
3448	Prefabricated metal buildings

APPENDIX D:
SIC CODE AND DESCRIPTION BY INDUSTRY SECTOR

Consumption (cont.)	3451	Screw machine products
	3452	Bolts, nuts, rivets, and washers
	3462	Iron and steel forgings
	3479	Metal coating and allied services
	3482	Small arms ammunition
	3484	Small arms
	3489	Ordnance and accessories, nec
	3491	Industrial valves
	3498	Fabricated pipe & fittings
	3523	Farm machinery and equipment
	3531	Construction machinery
	3537	Industrial trucks and tractors
	3541	Machine tools, metal cutting types
	3544	Special dies, tools, jigs & fixtures
	3559	Special industry machinery, nec
	3562	Ball and roller bearings
	3563	Air and gas compressors
	3585	Refrigeration and heating equipment
	3599	Industrial machinery, nec
	3612	Transformers, except electronic
	3679	Electronic components, nec
	3699	Electrical equipment & supplies, nec
	3714	Motor vehicle parts and accessories
	3715	Truck trailers
	3721	Aircraft
	3728	Aircraft parts and equipment, nec
	3731	Ship building and repairing
	3769	Space vehicle equipment, nec
	3842	Surgical appliances and supplies
	3965	Fasteners, buttons, needles, & pins
	4011	Railroads, line-haul operating
	4013	Switching and terminal services
	4142	Bus charter service, except local
	4212	Local trucking, without storage
	4213	Trucking, except local
	4214	Local trucking with storage
	4215	Courier services, except by air
	4221	Farm product warehousing and storage
	4222	Refrigerated warehousing and storage
	4225	General warehousing and storage
	4226	Special warehousing and storage, nec
	4231	Trucking terminal facilities
	4491	Marine cargo handling
	4493	Marinas
	4581	Airports, flying fields, & services
	9711	National security